STATE OF INDIANA

INDIANA UTILITY REGULATORY COMMISSION

In the Matter of the Petition of	)	
Indiana Bell Telephone Company,	)	
Incorporated d/b/a Ameritech Indiana	)	
Pursuant to I.C. 8-1-2-61 For a Three	)	Cause No. 41657
Phase Process For Commission	)	
Review of Various Submissions of	)	
Ameritech Indiana to Show Compliance	)	
with Section 271(c) of The Telecommunications	)	
Act of 1996	)	

MOTION TO SUBSTITUTE CORRECTED REPLY BRIEF ON EXCEPTIONS OF AMERITECH INDIANA

Petitioner, Indiana Bell Telephone Company, Incorporated d/b/a Ameritech Indiana ("Ameritech"), by counsel, respectfully requests that the Indiana Utility Regulatory Commission ("Commission") permit Ameritech to substitute a corrected Reply Brief On Exceptions Of Ameritech Indiana. In support of this request, Ameritech states:

1. The Reply Brief On Exceptions Of Ameritech Indiana refers to information from a monthly remedy report to rebut certain allegations made by AT&T regarding the amount of remedies paid in Illinois;

2. The report contains certain confidential data that, as a matter of course and procedure, are generally disclosed only to the applicable CLEC (and to the Commission upon request, under appropriate confidentiality provisions) to guard against the possibility that a competitor might glean or infer information about the CLEC's business activities and plans;

4. Ameritech inadvertently attached an excerpt from the report to its brief, as Exhibit 2;

5. By mistake, the last page of Exhibit 1 to the Reply Brief was left out of the filed copy of the Reply Brief On Exceptions Of Ameritech;

6. Ameritech seeks leave to substitute for the original brief a corrected version that excludes Exhibit 2, with the text of the brief revised accordingly, and including the last page of Exhibit 1;

7. Ameritech, by service to the parties hereto, asks all parties to destroy the original Reply Brief On Exceptions Of Ameritech Indiana.

WHEREFORE, Ameritech Indiana moves:

A. That the Commission permit the substitution of the Corrected Reply Brief On Exceptions Of Ameritech Indiana;

B. That the Commission remove, from the internet site maintained in this cause, the posted Reply Brief On Exceptions Of Ameritech Indiana, with all attachments; and

C. That the Commission grant such other and further relief as it may deem appropriate in the premises.

Respectfully submitted,

Bonnie K. Simmons (12067-49)
Ameritech Indiana
240 N. Meridian St., Room 1831
Indianapolis, IN 46204
Telephone: (317) 265-3676
Fax: (317) 265-3343
Attorney for Indiana Bell Telephone Company, Incorporated D/B/A Ameritech Indiana

CERTIFICATE OF SERVICE

I hereby certify that a true and correct copy of the foregoing was served upon the following this 11th day of April, 2001:

Karol Krohn
Office Of Utility Consumer Counselor
Indiana Government Center North
100 North Senate, Room N501
Indianapolis, Indiana 46204

Douglas W. Trabaris
Senior Attorney
At&T Corp.
222 West Adams Street; 15th Floor
Chicago, Illinois 60606

Michael J. Huston
Michael E. Allen
Baker & Daniels
300 North Meridian Street; Suite 2700
Indianapolis, Indiana 46204

Pam Sherwood Vice President
Regulatory Affairs, Midwest Division
Time Warner Communications
4625 W. 86th Street, Suite 500
Indianapolis, Indiana 46268

John Kern
2300 N. Barrington Road, Suite 400
Hoffman Estates, Il 60195

Frank Darr
National Regulatory Research Institute
1080 Carmack Road
Columbus, Oh 43210

Jack R. Boheim
President
Mtg Consulting
P.O. Box 2448
Mendocino, Ca 95460

Nikki Gray Shoultz
Sommer & Barnard, Pc
4000 Bank One Tower
111 Monument Circle
Indianapolis, Indiana 46204

Charles R. Mercer, Jr.
Sprint
One North Capitol Ave., Suite 540
Indianapolis, Indiana 46204

Robert K. Johnson
Christopher C. Earle
Bose Mckinney & Evans
2700 First Indiana Plaza
135 North Pennsylvania Street
Indianapolis, Indiana 46204

Ellyn Elise Crutcher
Associate General Counsel
McleodUSA Telecommunications
Services, Inc.
121 South 17th Street
Mattoon, Il 61920

Richard E. Aikman, Jr.
STEWART & IRWIN
Two Market Center
251 East Ohio Street, Suite 100
Indianapolis, In 46204

William Powers
111 Monument Circle, Suite 302
Indianapolis, IN 46204

Bonnie K. Simmons

STATE OF INDIANA

INDIANA UTILITY REGULATORY COMMISSION

In the Matter of the Petition of Indiana Bell Telephone Company, Incorporated d/b/a Ameritech Indiana Pursuant to I.C. 8-1-2-61 For a Three Phase Process For Commission Review of Various Submissions of Ameritech Indiana to Show Compliance with Section 271(c) of The Telecommunications Act of 1996	)))))))))	Cause No. 41657

CORRECTED REPLY BRIEF ON EXCEPTIONS OF AMERITECH INDIANA

Pursuant to the Commission's Order dated January 8, 2001, and January 24, 2001, Indiana Bell Telephone Company Inc. d/b/a Ameritech Indiana ("Ameritech"), by counsel, respectfully submits its reply brief on exceptions in the above-captioned proceeding.

INTRODUCTION

To read the briefs of the CLECs and Z-TEL, one would scarcely know that the FCC has thrice approved the very remedy plan that Ameritech has proposed and that they so bitterly oppose, or that the FCC has soundly rejected many of the arguments they offer in their opposition here. The most fundamental and egregious omission is Z-TEL's opening argument (at 2) that Ameritech's plan "fails to comport" with the FCC's five criteria for a reasonable plan. Whatever Z-TEL might want to say in its brief, the best judge of the FCC's criteria is the FCC that articulated them, and it has held the Ameritech plan *does* comport with the five criteria. *Texas 271 Order*, ¶¶ 420-429; *Kansas & Oklahoma 271 Order*, ¶¶ 270-278.

The opposition likewise founders when it tries to address the individual FCC criteria. Take the first: the FCC's holding that a remedy plan should establish "potential liability that provides a meaningful and significant incentive to comply with the designated performance standards." The CLECs contend that Ameritech's proposal "does not provide [Ameritech] with enough incentive . . . to provide the CLECs with the necessary service quality to allow competition to develop in Indiana." But the FCC reached precisely the opposite conclusion when it evaluated identical plans – and rejected identical CLEC objections -- in Texas, Kansas, and Oklahoma:

> We conclude that the total of $289 million in potential penalties placed at risk, on an annual basis, under the performance plans represents a meaningful incentive for SWBT to maintain a high level of performance. We thus disagree with commenters that suggest that this amount is insufficient and fails to provide adequate assurance of SWBT's compliance in the future. (*Texas 271 Order*, ¶ 424.)
>
> * * *
>
> We conclude that the total of $45 million for Kansas and $44 million for Oklahoma in potential penalties placed at risk, on an annual basis, under the performance plans represents a meaningful incentive for SWBT to maintain a high level of performance. As a percentage of the applicant's in-state net return, these penalty plans place the same amount at stake as the plans adopted in Texas and New York. We thus disagree with commenters that suggest that this amount is insufficient and fails to provide adequate assurance of SWBT's compliance in the future. (*Kansas & Oklahoma 271 Order*, ¶ 274.)

In the same vein, the CLECs call Ameritech's proposed cap on remedies under the plan an "absolute cap" that would "send the signal that once [Ameritech's] performance deteriorates . . . to a particular level (i.e. reaching the absolute cap) further deterioration is irrelevant." Wrong again, as the FCC demonstrated in Texas, Kansas and Oklahoma: "The performance plans adopted by the Texas Commission do not represent the only means of ensuring that SWBT continues to provide nondiscriminatory service to competing carriers. In addition to the $289 million at stake under this Plan, as noted above, SWBT faces other consequences if it fails to

sustain a high level of service to competing carriers." *Texas 271 Order*, ¶ 424. *Accord Kansas & Oklahoma 271 Order*, ¶ 274.

The opponents' inability to challenge the FCC's holdings leads them to criticize the proceedings that led to the FCC's grant of section 271 approval for Texas. The CLECs complain (at 4) that the Texas remedy plan, which served as the model for the one proposed by Ameritech here, "was crafted by SBC's attorneys and lobbyists and the Chairman of the Public Utility Commission of Texas in bilateral negotiations to which other parties . . . were excluded." Z-TEL similarly contends (at 1-2) that the Texas plan "was developed by SWBT and the Texas commission in meetings that were not open to the various [CLECs]." Once again, though, the FCC has squarely refuted the opponents' position. Contrary to their claim that the Texas remedy plan was developed in secret, the FCC found that "the Texas Commission ensured that its section 271 review process was open to participation by *all* interested parties" – a finding that the FCC gave "particular note" in agreeing with the Texas commission (and the Department of Justice) that Southwestern Bell had satisfied the conditions for long-distance entry. *Texas 271 Order*, ¶ 3 (emphasis added). And as we noted in our reply comments (at 9-10), many of the defining features of the so-called "secret" Texas plan (for example, the z-test, the use of a 95 percent confidence level, the use of a "k" table to address the remaining 5 percent error rate) were originated by CLECs.

That leaves the opponents to their truly last resort: the parochial assertion that "the Texas plan . . . is not suited to Indiana." CLEC Comments at 4. Ameritech disagrees: The underlying performance measures and standards that the proposed plan will enforce are virtually identical to those used in Texas. The CLECs, most of whom operate nationwide, agreed to import the Texas measures and standards here to promote consistency and uniformity – indeed, in several

instances in the collaborative the CLECs actively pushed for Texas measures and standards. Thus, it makes perfect sense that the associated remedy plan, which is designed to fit the underlying measures hand in glove, should also be similar to that in Texas. And in those cases where performance measures and standards for Indiana differ, Ameritech has tailored the plan accordingly.

The only specific argument the CLECs offer for deviating from the Texas plan in midstream -- their view that Indiana would have smaller sample sizes – simply adds more force to Ameritech's proposal. Ameritech's plan tackles small samples with statistical methods designed especially for small samples. The CLEC plan, by contrast, proposes the same methods for small samples that it does for large samples, even though statisticians -- including those employed by AT&T -- have found that "one-size-fits-all" approach unreasonable. Ameritech Exceptions at 15-16.

At any rate, the ultimate rejoinder to the CLECs comes from the FCC. It has approved the Texas plan for use not only in Texas but also in smaller states (Kansas and Oklahoma) and even in the Ameritech states (as a condition of its approval of the SBC/Ameritech merger). The opponents' forum-shopping is simply an admission that *no* jurisdiction has approved their proposals or bought their arguments, and is founded on their hope that if they continue to shut their eyes tightly enough to precedent and federal law, some commission somewhere will too. This Commission should not take the bait.

ARGUMENT

I. To Be Meaningful, A Remedy Plan Should Only Assess Remedies When Poor Performance Occurs.

One of the principal issues to be addressed in a remedy plan is to recognize that performance data are subject to random variation (particularly when they are sliced into small samples by month, CLEC, product, and geography) and to ensure that remedies are assessed only for real shortfalls in performance as opposed to random chance. Ameritech's plan meets that goal by using a two-step process: (1) use statistical analysis, which calls for remedies only when the amount of the performance shortfall is large enough that we can say, with 95 percent "confidence" that the shortfall is not due to random chance; and (2) use a "k" table to address the remaining 5 percent error rate. As we showed in our previous comments, the CLEC and Z-TEL plans fail to adequately address the problem of random variation, and are in fact designed to assess remedies where none are deserved.

A. Statisticians refer to two types of errors: A Type I error occurs when a statistical test falsely suggests failure (in this case disparity) based on random variation, where the reality was parity. A Type II error occurs when the test gives a passing grade (shows parity), where the reality was disparity.

Z-TEL barks (at 5) that Ameritech's plan "does absolutely nothing to balance Type 1 and Type 2 errors – nothing." Similarly, the CLECs contend (at 24) that "there is silence from [Ameritech] on whether balancing should be done." That is patently false. Ameritech's plan uses the same z-test, with the risk of Type I error set at 5 percent, that the FCC has deemed "a fair compromise" between Type I and Type II error (*New York 271 Order*, App. B., ¶ 17), and that AT&T itself has agreed "strikes a fair balance between the need to account for both Type I and Type II errors" (Ex. 1, at 13). As AT&T's expert explained (*id.*):

> If we choose to make the Type I error small, then the Type II error will be large; and conversely. AT&T proposes to set the Type I error at no more than the conventional level of 5%. This controls the frequency of false alarms to be at most 5% while making the probability of Type II errors small for violations that are of substantial size.

Likewise, the FCC expressly found that Ameritech's methodology "gives us a reasonable likelihood of detecting variations in performance not due to random chance [Type II errors], with few false conclusions that variations are not due to random chance [Type I errors]." *New York 271 Order*, App. B, ¶ 17.

It is instead Z-TEL that does "absolutely nothing to balance Type 1 and 2 error." Indeed, Z-TEL's plan makes no attempt to determine whether performance is truly poor, so there is no way to tell whether and how often it has assessed remedies correctly. All Z-TEL does is try to cram the distribution of wholesale data into arbitrary "zones" – based on retail performance in some unspecified past period that could have no connection or relevance to present circumstances – and punish Ameritech if the data doesn't fit the same pattern. There is absolutely no evidence that these zones have any scientific validity in any field, let alone that of statistics. Z-TEL even admits (at 7) that "Type I and Type II errors are not relevant" to its unscientific plan.

B. Meanwhile, the CLEC plan claims to balance Type I and Type II errors, but does not. As with any other test, to decide whether a statistical test yields the right answer or an error, one needs to define what the right answer is. If a test suggests disparity, you need to know what parity is to find out if there has been a Type I error. That is simple, because "parity" or equality is a straightforward concept that is easy to define and agree upon. By contrast, if a test suggests parity, the only way to know if there has been a Type II error is to define *dis*parity. And there's the rub. A difference in performance may exist, but it may be so slight that it has no impact on

customers or the market. Thus, defining disparity means making some critical assumptions about what difference in service provided to retail and wholesale customers would have a material impact on competition. In order to make an informed decision about what affects competition, one would need a vast array of information about how customers react to differences in levels of service. At this time, such information does not exist, and it certainly does not exist for each performance measure, local telecommunications product, or geographical disaggregation, let alone for every relevant combination that is to be included in the remedy plan. As the FCC found in its *New York 271 Order* (¶ 46):

> Finally, we note that a determination of whether the statutory standard is met is ultimately a judgment we must make based on our expertise in promoting competition in local markets and in telecommunications regulation generally. We have not established, nor do we believe it appropriate to establish, specific objective criteria for what constitutes "substantially the same time and manner" or a "meaningful opportunity to compete." We look at each application on a case-by-case basis and consider the totality of the circumstances, including the origin and quality of the information before us, to determine whether the nondiscrimination requirements of the Act are met. Whether this legal standard is met can only be decided based on an analysis of specific facts and circumstances.

Striking an *exact* balance between the rate of Type I and Type II errors (as opposed to the reasonable balance provided by Ameritech's plan) would first require an accurate definition of a Type II error, which in turn would require an accurate definition of "disparity." The CLEC plan would thus necessitate a multitude of complex analyses that are sure to produce an equal number of disagreements as to the appropriate definition of disparity for each measure, product, and geographic region. And in the end, there would be no assurance that the end result, fraught as it would be with subjectivity, would be accurate.

Unable to solve the critical first step of defining disparity, the CLECs try to simply skip over it. They simply assume that a difference of 0.25 times the population standard deviation (a

statistical way to measure the degree of variation in a set of data) would constitute a material disparity or "delta" to use the CLECs' nomenclature. A fixed definition of disparity is arbitrary; there is no reason to presume that a difference of .25 times the population standard deviation would affect competition or even be noticed by consumers. And the CLECs provide absolutely *no* evidence to support their assumption, even though it is the starting point of their entire balancing methodology. (In fact, the BellSouth plan for Louisiana, which the CLECs claim provides support for their proposal, sets "delta" at a significantly larger level of a full standard deviation – an approach that yields results that, for most sample sizes, correspond more closely to the Ameritech plan than to the CLEC plan. In addition, the Louisiana plan has a number of other features that would greatly reduce the number of Type I errors produced. The CLEC proposal here omits these features even though they cite the Louisiana plan for support. .)

More fundamentally, the underlying approach – using a fixed number instead of judgment and expertise to define disparity – is directly contrary to the FCC's holding: "We have not established, *nor do we believe it appropriate to establish, specific objective criteria* for what constitutes 'substantially the same time and manner' or a 'meaningful opportunity to compete.' * * * Whether this legal standard is met can only be decided based on an analysis of specific facts and circumstances." *New York 271 Order*, ¶ 46. *See also Kansas & Oklahoma 271 Order*, ¶ 29 (reiterating the need to consider each application on a case-by-case basis and to analyze the totality of the circumstances). Indeed, the CLECs' own remedy plan agrees that using a fixed value to establish materiality would "not seem sensible." Joint Comments of the Indiana CLECs, at 31 (filed Feb. 9, 2001).

In short, Ameritech's plan has been found – by the FCC and by AT&T – to strike a reasonable balance between Type I and Type II errors, because it "controls the frequency of false

alarms . . . while making the probability of Type II errors small for violations that are of substantial size." Ex. 1, at 13. Striking an *exact* balance is impossible, because it requires establishing (and then constantly updating) an exact definition of disparity for every single one of the thousands of performance measures and categories. Striking an *arbitrary* balance – by merely setting an arbitrary definition of disparity without any evidence to support it, the way the CLECs propose – virtually guarantees that Ameritech will be punished solely on the basis of random chance, and that remedy payments will have little connection to real performance. In fact, as we stated in the Exceptions, we estimate that the CLEC plan would penalize Ameritech approximately $8 million per month even if it provided perfectly non-discriminatory service.

C. Even more offensive is Z-TEL's charge (at 7) that Ameritech's method for addressing random variation in performance data that are assessed against fixed benchmarks does not have "statistical merit." The Z-TEL and CLEC plans do not perform *any* analysis to address random variations in performance on these measures; to use Z-TEL's words (at 7), *theirs* is the approach that "violates a number of basic laws of mathematics and statistics." And it is particularly galling that Z-TEL should profess to wave the banner of statistical purity here – after all, Z-TEL openly boasts that its entire plan has no foundation in statistical science or principle. All Z-TEL is trying to do is cloud the issue. The bottom line is (1) Ameritech's methodology for testing benchmark measures, like the other features of its plan, was developed as a compromise in the Texas collaborative proceedings, (2) it has been approved by several state commissions and the FCC, and (3) in the collaborative proceedings here, Ameritech was perfectly willing to discuss modifications to its methodology that had any scientific merit, but the opponents abjectly refused to discuss any method for addressing random variation in these measures.

D. As described in its previous comments, the statistical methodology of Ameritech's plan is designed to assess remedies with 95 percent confidence: in other words, on average it would assess remedies 5 percent of the time even if performance were perfectly nondiscriminatory, based solely on random chance. Thus, Ameritech's plan includes a "k" table, which is designed to exclude remedies for those first few perceived shortfalls in performance that one would expect are due to random chance, given the 5 percent "Type I" error rate. The number of excluded measures, "k," is based on the total number of measures and categories tested for each CLEC, and Ameritech's plan includes a table that gives the number of exclusions. (Ameritech's plan takes a conservative approach excluding, first, remedies on measures that have lower priorities or fewer occurrences.)

Z-TEL argues that the values of "k" are too high because in some cases "k" exceeds 5 percent of the measures tested. Z-TEL's argument is based on its simplistic assumption that "k" should always be 5 percent, because it is designed to address the Type I error rate of 5 percent. But saying that a test yields Type I errors 5 percent of the time *on average* is the same as saying that the rate of flipping a fair coin and getting "heads" is 50 percent on average. For smaller sample sizes (say 100 flips), the actual number of heads will vary, and could be higher or lower than 50 percent; we would expect that about half the time it will be higher and half the time lower. Similarly, given a Type I error rate of 5 percent, the number of false alarms for a given CLEC will likely exceed 5 percent half the time, and remedies would be paid to one of every two CLECs even when overall service is nondiscriminatory. Thus, setting "k" at 5 percent is not statistically valid. The proper value of k will account for the fact that a 5 percent Type I error rate will not always yield exactly 5 percent false alarms.

The method for getting the valid number of k is based on a standard mathematical formula known as the "binomial distribution." As one would expect, where the number of measurement categories is small, the number of measures excluded is slightly higher than 5 percent: hence, 8 categories would be excluded if 100 categories had data. But as the number of categories increases – and there are now several thousand measurement categories - the value of k does approach 5 percent.

But the Commission need not just take *Ameritech's* word that the binomial formula is scientifically valid. The k table was promulgated by AT&T (Ex. 1, at 28) in the FCC's docket on performance, and it was proposed by MCI in the Texas collaborative that led to this plan. As the Commission will note, the value of k in AT&T's table (just as in Ameritech's k table) is higher than 5 percent of the number of tests, and approaches 5 percent as the number of tests increases.

Notably, neither one of these CLECs joins Z-TEL's objection to the formula here. Instead, they object to the outcome, claiming that it "eliminated" 95% of the dollars allegedly due AT&T in Illinois for the month of December 2000. But they provide no context as to how the plan reached that result. The omitted facts tell the real story:

1. Ameritech passed the performance test on over *93 percent* of those categories and missed on less than 7 percent.

2. Given that we would expect a 5 percent failure rate due solely to random variation, Ameritech's k-table excluded most of the misses.

Thus, the reason AT&T did not get all the remedy dollars it wanted was not due to any deficiency in Ameritech's plan, but because wholesale performance was stellar. That result is not "obscene" as AT&T describes it, but entirely consistent with the FCC's statement that remedies should be assessed only when poor performance occurs.

II. To Be Meaningful, The Remedy Amount Should Reflect The Impact Of The Performance Measure And The Volume Of Customers Affected.

Ameritech's plan computes remedies by multiplying a "base remedy" amount to reflect the following factors:

- *severity*: the difference between actual performance and adequate performance;
- *volume*: the number of occurrences (that is, the number of CLEC transactions involved), or a flat per-measure factor if the number of transactions does not reflect volume;
- *duration*: the remedy amount escalates if poor performance continues for more than a month; and
- *priority*: the impact of the measure that experienced a performance shortfall.

There is little real disagreement over the first factor: that of severity.

A. With respect to volume, Z-TEL complains that per-occurrence remedies "discriminate against entry strategies that by their nature have smaller sample sizes." And Z-TEL is right – but that is precisely why Ameritech does *not* use per-occurrence remedies for *all* measures, and why Ameritech instead uses per-measure remedies for those measures "that by their nature have smaller sample sizes" -- i.e. where a single transaction, like local switching, affects numerous customers. Further, Ameritech's plan provides for trebling Tier 2 remedies on measures related to nascent services (like DSL loops, ISDN loops, and resold ISDN) that currently run low volumes.

What the CLECs and Z-TEL fail to see is the other side of the coin: per-measure remedies (on which their plans depend) discriminate against entry strategies and CLECs that support *larger* volumes and sample sizes, and reward CLECs that do only token business. If a CLEC that has one customer receives the same remedy as the CLEC that has a million, the remedy plan creates no incentive to compete. Applying per-measure remedies to *all* performance measures, even those that support large volumes, encourages CLECs not to have an "entry strategy" – or even a desire to enter -- at all.

B. As for duration, the opponents try to muddle the record and then claim that Ameritech's methodology is too complicated. But in reality Ameritech's plan is quite simple. If performance warrants a remedy on a given measure in one month, a remedy is assessed. If poor performance continues for that measure in the next month, another remedy is assessed, this time at a higher rate, and so on through the six month. The CLECs insinuate that there is a "ceiling" on remedies after that, but if poor performance persists, the remedies do not cease. They continue to be assessed, at the maximum rate. True, the base amount of each monthly remedy does not continue to increase. There must be some maximum monthly amount before the remedies on a single measure reach absurd levels and that single measure dwarfs the others, motivating Ameritech to focus on one measure and entry strategy to the exclusion of others. But Ameritech would still pay remedies at a high level (even higher, because they may be multiplied by the number of occurrences in each month), and because those remedies are not exclusive of others the CLEC or the State or the FCC would have, poor performance that really has a competitive impact would surely be redressed.

What happens if poor performance does not persist – in other words, what if Ameritech solves the underlying problem? Under the Z-TEL plan, not much: the remedy amount for that

measure remains at whatever high level it reached before Ameritech's solution, so if there is poor performance in any future month, no matter how distant from or unrelated to the original problem, it exacts a high price. Again, that reduces the incentive to improve and makes Ameritech focus unduly on some measures to the detriment of others. Even the CLECs do not take this approach. Under Ameritech's plan (and even under the CLEC plan), the duration factor is reset to the starting point, to reflect and reward improved performance.

But Z-TEL's plan doesn't just fail to reward improvement – it uses improved performance in one month to raise the bar for future months, in effect *punishing* Ameritech for improvement. The following example illustrates the improper incentives Z-TEL's plan would create. Assume that Ameritech not only provides nondiscriminatory service, but also makes wholesale repairs 30 minutes faster than retail. In January, retail repairs take 6 hours, while wholesale take 5.5 hours. A mild February brings the retail interval down to 3.5 hours, with wholesale only 3 hours. March and April bring more normal conditions, and in each month the retail interval stabilizes at 4.5 hours, while wholesale remains at 4.

Does that performance warrant punishment? Certainly not. Wholesale performance was consistently better than retail, and performance for wholesale and retail operations improved as the year went on. But Z-TEL's plan would impose punishment nonetheless. It would set the standard for future performance using February's retail performance of 3.5 hours, even if that performance was not representative of current conditions. Thus, Ameritech would be punished in March and April for wholesale performance of 4 hours – even though wholesale performance was consistently faster than retail performance in those months. Worse, the penalties would continue to escalate. The problem, of course, is not with Ameritech, but with Z-TEL's improper attempt to compare current wholesale performance to *past* retail performance.

Given the reasonable incentive structure, on which Ameritech's duration factors are based, it should come as no surprise that the FCC saw through and rebuffed the same arguments the CLECs try to rehash here:

> The structural elements of the Plan appear reasonably designed to detect and sanction poor performance when it occurs. Commenters have raised specific criticisms, arguing, for example, that the Plan fails to deter targeted discrimination directed against individual competing carriers, and does not include penalties that escalate with the magnitude or duration of the performance shortfall. These criticisms do not undermine our overall conclusion that the Plan provides a meaningful incentive to provide nondiscriminatory performance in the future. We find it significant that the Texas Commission considered and rejected most of these arguments.

C. We have already anticipated the CLECs' argument that Ameritech's priority weights for measures – which apply equally to similar measures regardless of product, service, or entry strategy – favor some entry strategies over others. And we have previously shown that the CLECs' assertions have no basis. At this point, we add only that Ameritech afforded the CLECs every opportunity to suggest changes to the priority weights to address any real concern about inequality they might have had – and they refused to negotiate the matter at all. Rather, their goal is to assign every measure the same high priority, regardless of its importance and in spite of the fact that many measures closely track other, related measures (*e.g.* the rate of missed due dates in total tends to follow the rate of missed due dates due to lack of facilities), such that giving both measures the same priority would be unfair double-counting.

III. Ameritech's Plan Provides A Reasonable And Practical Structure for Testing Performance.

Ameritech's plan is self-executing: It provides for Tier 1 remedies to be credited automatically against the applicable CLEC's monthly bill. The use of such credits is sensible. The amount a CLEC owes for Ameritech's performance of interconnection agreement obligations, and the remedy the CLEC is entitled to receive for any poor performance by Ameritech, are naturally related. They should be combined and offset the same way that service credits to end users are offset against the end users' bills. Credits are also practical. Unlike checks – which have to be processed, approved, signed and mailed by Ameritech, received and cashed by the CLEC, and then cleared by the banks – bill credits have an immediate financial benefit.

Thus, Ameritech is baffled that the CLECs and Z-TEL have both directed such vociferous opposition to this unobjectionable feature of the Ameritech plan. None of their objections has merit. We can first discard the opponents' hypothetical, in which a CLEC stops doing business in Indiana and might have no monthly bill to be credited. Putting aside the question whether the plan or the Commission should go to any extra trouble to accommodate CLECs that abandon the state, and putting aside the fact that a CLEC is likely to have *some* leftover billing to credit when it does leave, Ameritech's methods and procedures provide that such a CLEC can request a check under the Ameritech plan (if it has a net credit balance), just as it would under the competing remedy plan (and just as end users do if they terminate phone service before service credits are issued).

The opponents' arguments that credits are not sufficiently visible to them or to Ameritech management are equally unfounded. For CLECs, credits under the Ameritech plan would be identified as a separate line item titled "Liquidated Damages" on the carrier bill. Ameritech also

publishes a monthly "Remedies Report" on its CLEC website, which tells each CLEC the performance measures for which remedies were assessed, the remedy amount, and any assessments excluded by means of the k table described above. This gives the CLEC ample information to understand and check their bill.

Credits under the plan would also be quite visible to Ameritech management. Ameritech assigns ownership of performance measures (and the associated remedies) to senior executives in charge of the operating unit that is responsible for the function measured; for example, remedies for missed due dates for unbundled loops are assigned to the Network President, while remedies related to firm order confirmations are assigned to the President of Industry Markets, the organization that runs Ameritech's Local Service Center. Remedy credits for those measures are charged directly to the operating expense of the organization and manager that owns them – and can thus impact future compensation, which is based in part on meeting the operating budget goal. We believe that such consequences are more visible and meaningful to managers than signing a check. At any rate, Ameritech's plan also provides for Tier 2 remedies (paid to the State), which *are* issued by check, so it still has whatever impact the act of signing of a check would have.

Finally, credits would not increase the overall administrative cost of the plan. The systems and procedures for issuing credits are already in place and in use in Texas, Kansas, Oklahoma, Illinois and Ohio. It would cost more – for Ameritech and the CLECs -- to develop new procedures for processing checks (and to incur the costs of check processing and delivery) than it would to simply carry over the existing procedures for bill credits and use the existing bills to process credits.

IV. The Commission Should Reject The CLEC Proposal To Replace The Nondiscrimination Standard with "Parity with a Floor."

It is beyond dispute that one of the cornerstones of the 1996 Act is nondiscrimination or "parity": the principle that an incumbent LEC should give its competitors the same treatment it gives itself. The Act uses the word nondiscrimination too many times, in too many places, to contend otherwise. It is equally well established that an incumbent LEC cannot be required to treat competitors *better* than itself. The Eighth Circuit (vested with exclusive authority under the Hobbs Act to review the FCC's rules implementing the Act) has said so, twice, and has twice struck down rules that would have required superior quality. Nevertheless, the CLECs propose to set fixed benchmarks for performance and to assess remedies if Ameritech fails to meet those benchmarks – even if Ameritech is treating CLECs the same or even better than itself. The CLECs make no attempt to hide that that is exactly what they contend, and they make no attempt to address the Eighth Circuit's controlling decisions.

What is the basis for the CLECs' frontal assault on the Act? In reality, there is none. The bulk of the CLECs' argument is neither here nor there: They complain that they are not now receiving parity treatment. Ameritech disagrees, and notes that the CLECs produce nothing but conclusory complaints without hard evidence. But that is beside the point: Saying there is not enough parity only proves the need for a parity standard, and there is no dispute that Ameritech's plan offers just that. What the CLECs are trying to do is impose a standard of *superior* treatment. To support that standard all they can offer is the assertion that parity is not good enough for them, because they lack the resources and reputation to compete on an equal footing with Ameritech. Coming as it does from long-established, mega-billion, multi-national corporations like AT&T and MCI/WorldCom, that assertion is laughable. More fundamentally, the CLEC argument should be addressed to Congress, not the Commission. Congress wrote the

law based on nondiscrimination and only Congress can rewrite it. The purpose of this proceeding and of the remedy plan is to enforce the Act, not amend it.

V. The Commission Should Ignore Z-Tel's Untimely Revisions To Its Plan

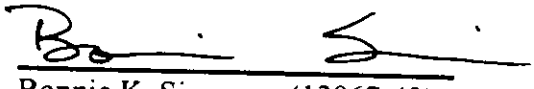
The Commission described its principles for effective remedy plans on November 9, 2000, and set February 9, 2001, as the deadline for parties to submit proposed remedy plans that would, in their view, meet the Commission's principles. The parties submitted their plans on that data, and commented on the principles for remedy plans set forth by the Commission. But in its March 8 comments – one month after the deadline for plan submission – Z-TEL tried to submit a new plan, which it described as "a revised Zone-Parity proposal which Z-TEL contends would satisfy all of the Commission's principles outlined in its November 9, 2000 docket entry."

Z-TEL made no attempt to tell the Commission what it had revised or why. The revised plan contains but a single "red-line" mark, on page 30, which might lead the reader to believe that Z-TEL changed only the title of one chapter heading. That impression would be wrong. In fact, Z-TEL changed a whole lot more: It developed an entire methodology (different from, but just as unscientific as the rest of its plan) for analyzing and computing remedies on measures that are expressed as percentages: for example, the percentage of due dates missed, or the percentage of firm order confirmations issued within 2 hours. This was apparently done in response to Ameritech's observation that Z-TEL's original plan did not even mention such measures, which constitute the bulk of the measures approved by the Commission.

The Commission should strike Z-TEL's untimely submission. The parties had ample opportunity to develop, explore, and revise plans during collaborative proceedings that stretched over 6 months – both here and throughout the region – and the Commission gave them three

months to address its principles. Z-TEL never asked for more time; in fact, even now Z-TEL does not say that the time allotted was insufficient. By trying to sneak a new plan through now - after the parties submitted two rounds of comments - Z-TEL has admitted that the plan it submitted on February 9 was deficient, and that its representation that the plan satisfied the Commission's principles was false. The Commission should consider that failure when it reads Z-TEL's latest representation that its plan "would satisfy all the Commission's principles."

Respectfully submitted,



Bonnie K. Simmons (12067-49)
Ameritech Indiana
240 N. Meridian St., Room 1831
Indianapolis, IN 46204
Telephone: (317) 265-3676
Fax: (317) 265-3343

Attorney for Indiana Bell Telephone Company, Incorporated D/B/A Ameritech Indiana

Before the
FEDERAL COMMUNICATIONS COMMISSION
Washington, D.C. 20554

In the Matter of	)	
	)	
Performance Measurements and	)	
Reporting Requirements	)	CC Docket No. 98-56
for Operations Support Systems,	)	RM 9101
Interconnection, and Operator	)	
Services and Directory	)	
Assistance	)	

Affidavit of Dr. Colin L. Mallows

Colin L. Mallows, being duly sworn, deposes and says:

1. I am a Technology Consultant at AT&T Laboratories. I make this affidavit in support of AT&T's comments regarding the use of statistical methods to determine whether incumbent local exchange carriers ("ILECs") are providing nondiscriminatory, i.e., parity, service to competing carriers ("CLECs"). I understand this is a requirement of law under Section 251 of the Telecommunications Act of 1996 ("Act").

Qualifications

2. I have been a professional statistician for nearly 45 years. I obtained a B.Sc. in Mathematics in 1951 and a Ph.D. in Statistics in 1953, both from University College, London. After two years in the British Army I became a lecturer at University College in the area of statistics. Since 1960, I have been employed at AT&T (formerly Bell

Laboratories, becoming Head of the Statistical Models and Methods Research Department in 1969. I relinquished that title in 1986. From 1960 through 1964, I was also an adjunct associate professor at Columbia University, teaching courses in statistical analysis.

3. I am a Fellow of the American Statistical Association ("ASA"), and I served as an associate editor of Journal of the American Statistical Association from 1966 to 1971, and again from 1986-1989. I am also a Fellow of the Institute of Mathematical Statistics ("IMS"), and an elected member of the International Statistical Institute. I was twice elected to the Council of IMS, and have served on various committees of the IMS and ASA. In 1997 I was honored by being named Fisher Lecturer at the Joint Statistical Meetings held by the ASA, IMS, the International Biometric Society and the Statistical Society of Canada.

4. I have published over 100 papers, with a large number of co-authors, in a variety of journals. My name is attached to several well-known statistical techniques, including the Cp-plot for selecting regression variables, the phi-model for analysis of ranking data, and a weighting scheme for robust linear regression. My professional interests include foundations, data analysis, statistical graphics, time series, robustness, software reliability,

moment-problems and Chebychev inequalities, combinatorics and coding theory.

Introduction

5. I have reviewed the Commission's Notice of Proposed Rulemaking ("Notice") in this proceeding, focusing on its discussion of the use of statistical analysis as a means of determining whether ILECs are providing parity service to new competitors. The Notice (¶ 34) is clearly correct that "reporting averages of performance measurements alone, without further analysis, may not reveal whether there are underlying differences in the way incumbent LECs treat their own retail operations in relation to the way they treat competing carriers." Thus, it properly proposes to require the use of statistical tests to determine whether measured differences in average ILEC performance for themselves and competitors "represent true differences in behavior rather than random chance."

6. As the Commission is aware, AT&T has supported the use of statistical tests to determine whether an ILEC has met its statutory obligations. Earlier this year, AT&T provided the Commission with a concept for applying statistical analysis to ILEC performance measurements.[1] The AT&T Statistical Ex Parte provided a methodology, given the

---

[1] Ex parte letter from Frank S. Simone, AT&T to Magalie Roman Salas, FCC, CC Docket No. 96-98, RM9101, dated February 3, 1998 ("AT&T Statistical Ex Parte").

presence of random error, to determine if an ILEC has complied with its statutory obligations when it reports results of numerous individual parity measurements, some of which show "worse" results for CLECs than for the ILEC.[2]

7. AT&T's Statistical Ex Parte correctly recognized that each of the individual tests of ILEC performance contained statistical Type I error. Thus, it is appropriate to use a Type I error concept when reviewing the ILEC's parity tests in the aggregate to determine whether the ILEC has met its nondiscrimination obligations. AT&T's Statistical Ex Parte thus described the use of a three-part analysis to determine whether ILEC measurements and reported results, when viewed in the aggregate, represent nondiscriminatory performance.[3]

8. Since that time, I have been asked to review and comment upon AT&T's Statistical Ex Parte and provide additional insight on the use of statistical tests in this

---

[2] Since most of the measurements for these purposes are measurements of time, a "worse" result for a CLEC is usually a larger value, e.g., a 5-day installation interval for a CLEC is worse than a 3-day interval for the ILEC.

[3] AT&T's proposal recommended establishment of separate thresholds for: (1) the maximum number of "failures" on a monthly report that could reasonably represent mere randomness resulting from the measurement process rather than disparity of performance; (2) repeated failures on specific performance measurements in consecutive months; and (3) measurements showing extreme differences in average performance for the ILEC and CLECs. Id., p. 3.

context. As described in Section I below, the more detailed statistical methodology that is proposed here requires only a two-part analysis and provides the ILECs with more leeway than the original AT&T proposal. Nevertheless, I believe that it provides a valid statistical comparison of the ILECs' actual performance for itself and CLECs.

Summary of Testimony

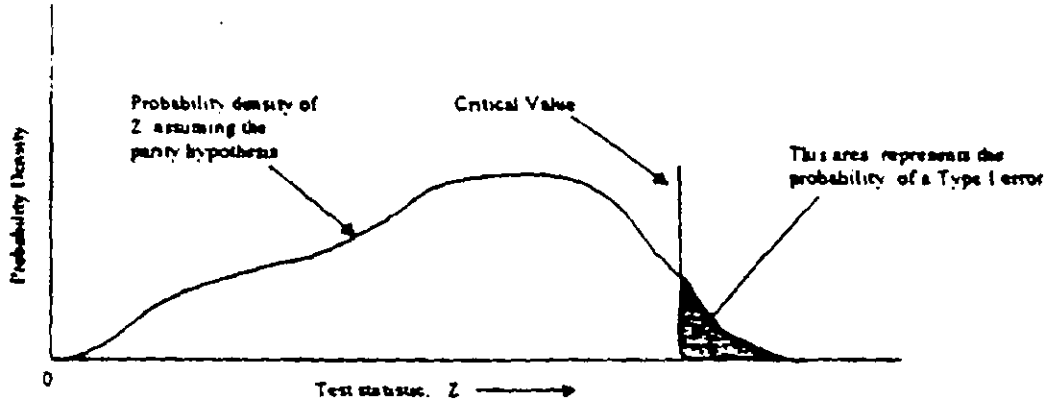
9. Specifically, my testimony below shows that AT&T's proposed methodology satisfies the Commission's desire to assure that reported differences in ILEC performance are statistically meaningful. With respect to individual tests of ILEC performance, there are three key components in developing an appropriate statistical methodology. First, the modified z-statistic proposed by LCUG provides an appropriate test statistic to determine whether there are significant differences in the mean and the variance of an ILEC's performance for itself and for CLECs. Second, a one-tailed test with Type I error held at the 5% level strikes a fair balance between the need to account for both Type I and Type II errors. Third, the t-distribution provides a useful basis for calculating the critical value for individual tests of ILEC performance, which is used to determine whether CLECs have been given equal treatment by the ILEC. Moreover, in those cases where the sizes of the ILEC and

measurements, from which the observed measurements are assumed to be drawn. We cannot observe these populations, and must base our test procedures on the observed samples. If the null hypothesis is accepted through the use of the chosen tests, then any differences in the ILEC's performance results for itself and the CLEC are deemed "statistically insignificant," and parity can be assumed.

13. All such statistical tests have three components. First, the test designer must select a test statistic, which is a formula that produces a single number summarizing the observed ILEC and CLEC data. Next, an acceptable Type I error probability must be adopted. The error probability represents the test designer's tolerance for falsely rejecting parity when it exists (Type I error is discussed in Section I.B below). Finally, the test designer must derive, from probability theory or known data, the probability distribution of the test statistic, describing the variability of performance under the null hypothesis.

14. Once these components are established, the test designer can determine (usually from a statistical table) a "critical value" against which to compare the computed value of the test statistic that is based on the actual results. If the test statistic is less than the critical value, it can be inferred that the ILEC's performance has "passed" the test of parity. If, however, the computed test statistic is

greater than the critical value, the ILEC's performance is judged to be not at parity, and the ILEC has "failed" the parity test for that measurement. The relationship between the performance distribution under the null hypothesis and the critical value is demonstrated graphically below.



A. **Test Statistic: The Commission Should Use The Modified Z-Statistic Recommended By LCUG.**

15. The modified z-statistic recommended by the Local Competition Users Group ("LCUG") is an excellent choice of test statistic in these circumstances. The "z-statistic" is a standard test statistic.[4] It is used to determine if the

[4] The formula for the z-statistic (also called the t-statistic), for the case where the observations are of measurements rather than proportions or rates, is

$$z = \frac{(\bar{Y} - \bar{X})}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)S^2}}$$

average results (or means) drawn from two separate performance samples (here the monthly ILEC performance data for itself and CLECs) have population means that are equal.[3] Thus, the standard z-statistic formula can determine whether, based on the reported results, the ILEC's average performance for itself and for CLECs is the same.

16. However, it is not enough to test for a difference in means alone. In order to obtain parity, CLECs are entitled to service from the ILEC that produces both the same mean performance and also the same variance in performance.[4] The z-statistic, in its standard form, is not

---

where $\bar{X}$ (resp. $\bar{Y}$) is the average of the ILEC (resp. CLEC) measurements, m (resp. n) is the number of these measurements, and S is a measure of the scale of variation of these measurements. The usual situation is that the statistical test is designed to detect a difference in the population means of the ILEC and CLEC measurements, assuming the population variances to be equal. In this case the standard choice for S^2 is

$$S^2 = S^2_{pooled} = \frac{(m-1)S^2_{ILEC} + (n-1)S^2_{CLEC}}{m+n-2}$$

[3] Similar statistics can also be used to detect differences in proportions and rates.

[4] The Commission also recognizes that it would be discriminatory if the ILEC has the same mean performance time for itself and CLECs but the variability of its performance for CLECs is greater (see Notice, Appx. B, ¶ 4 ("variability of response times . . . may affect the competitiveness of a competing carrier but may not be reflected in a comparison of average response times")). For example, CLECs would be at a commercial disadvantage if ILEC retail customers could always rely on an installation period of 4 days while installation dates for CLECs ranged from 2-6 days or more.

designed to detect differences in variance between CLEC and ILEC performance.

17. In order to create a single test that can account for both of these factors, LCUG proposes a modification that will make the statistical test have the power to detect whether the ILEC's variance in its performance for CLECs is greater than the variance in its performance for itself. Specifically, LCUG proposes to use the ILEC variance, rather than the "pooled" variance, in calculating the z-statistic.[7] This proposal is based on well-supported statistical testing principles and combines the power of tests of means and tests of variance. Thus, if the test proposed by LCUG is used, there would be no need to develop a separate test of the equality of variances.[8]

18. Use of the LCUG modified z-statistic, rather than the more conventional form that uses a "pooled" variance, is appropriate here because the problem here is different from

---

[7] The LCUG proposal is to use $S^2=S^2_{ILEC}$. The resulting test statistic has the same distribution theory as the conventional one (using S^2_{pooled}) except for changing the "degrees of freedom" from m+n-2 to m-1. The effect of this change will be small if the parity hypothesis holds, since as the incumbent monopolist, the ILEC sample is likely to be much larger that the CLEC sample.

[8] See Notice, Appx. B, ¶ 4. It should also be noted that the use of separate tests for differences in averages and differences in variance would reduce the power of each seperate test. Thus, it is preferable to use a single test that is sensitive to cases where both the mean and variance can increase.

that addressed in the standard texts. In the standard development, it is assumed that if the null hypothesis fails, it is only because the population means are different; the population variances are assumed to stay equal. This assumption is not appropriate here, because an increase in the CLEC variance would be a violation of parity, and the test should be able to detect it.[9]

19. As described above, the denominator of the formula for the z-statistic requires use of a figure for variance. Contrary to the suggestion of some ILECs,[10] the most appropriate variance to use in this case is the variance of the ILEC's performance for itself during the reporting period. This sample variance is the best available estimate of the variance of the ILEC process. Moreover, the entire purpose of the examination is to determine whether the ILEC is providing CLECs at least the same level of service as it provides to itself and its retail customers. Thus, for this

---

[9] Another standard form of the z-statistic is designed for the case where the two population variances may differ even under the null hypothesis. In this case one replaces $\left(\frac{1}{m}+\frac{1}{n}\right)S^2$ by $\frac{S^2_{ILEC}}{m}+\frac{S^2_{CLEC}}{n}$ This form of the statistic is inappropriate here since under the parity hypothesis the two population variances are equal. Use of this form would reduce the probability of detecting violations of parity.

[10] I am informed that some BOCs have suggested that the variance used in the formula should be based solely on the variance experienced by the CLECs, and others have suggested the use of the pooled variance.

purpose, variance in the ILEC's performance is the standard against which the performance for CLECs should be measured.

B. The Error Probability Should Be Based On A One-Tailed Test With Type I Error At No More Than The 5% Level.

20. In determining an appropriate Type I error probability for the statistical test, it is important to recognize that any probability rate above 0% means that the statistical test will produce errors.[11] It is also important to understand that there are two distinct types of testing errors. "Type I" errors occur when a statistical test shows that two sets of results (here for the ILEC and CLEC) are inconsistent with the null hypothesis (i.e., are not in parity) when in fact the null hypothesis is true. "Type II" errors are the opposite. They occur when a statistical test indicates that the outcomes are in parity, but parity does not in fact exist. Both types of errors are possible.

21. There are two "tails" to Type I errors, but the Notice (Appx. B, n.3) correctly notes that only one is pertinent here: errors relating to cases in which the ILEC's performance for CLECs is worse than its performance for itself. Under the Commission's rules, CLECs are entitled to performance that is "at least equal" to the performance the ILEC provides to itself. Those rules are not concerned with

---

[11] AT&T Statistical Ex Parte, p. B-1.

cases where, unintentionally, the ILEC provides CLECs with a level of performance that is better than the performance it provides to itself.[11] Thus, the Commission's rules themselves argue for a one-tailed test.

22. It should also be recognized that Type II errors are as real as Type I errors. Thus, there may be cases in which the ILEC is not in fact providing equal service to CLECs, but purely by chance the statistical test fails to reject the parity hypothesis. Thus, it is necessary to strike a balance between the two types of errors. If we choose to make the Type I error small, then the Type II error will be large; and conversely. AT&T proposes to set the Type I error at no more than the conventional level of 5%. This controls the frequency of false alarms to be at most 5% while making the probability of Type II errors small for violations that are of substantial size. Using a one-tailed test for Type I error at about the 5% level thus strikes a reasonable balance.[12]

---

[11] I am also informed that CLECs are not entitled to demand performance better than the ILEC provides to itself. Thus, there is no reason to believe that ILECs would intentionally provide their competitors with a higher grade of service than they provide to themselves and their retail customers.

[12] For general information supporting the 5% level, see AT&T Statistical Ex Parte, pp. B-1-B-2.

C. **Probability Distribution Should Be Based On The T Distribution Or A Permutation Distribution Analysis.**

23. For moderate or large sample sizes, it is appropriate to use the Student t (or "t") distribution to determine the critical value for the test. Use of this distribution, which is readily available in table form, is simple and straightforward and will produce statistically reliable results.

24. The published tables of critical values, using the t-distribution, are based on the assumption that the two populations (of ILEC and CLEC measurements) are exactly Normal. In practice, we will not have Normal distributions, and so these critical values are only approximations. There has been much debate as to the minimum sample sizes for which the tabulated values become acceptable approximations: numbers such as 10 or 30 have been suggested. But this must depend on the shape of the probability density function[14] of the populations, because there exist populations for which the approximation will never be adequate, even for very large sample sizes. In advance of reviewing the actual data, it is impossible to say for what size samples the tabulated values will be acceptable. Nevertheless, assuming that very large values of the observations do not occur and the populations have approximately symmetrical probability

[14] See the graph in ¶14 of these comments for an example of a probability density function.

density functions, I would guess that the tabulated values would be acceptable, provided that both the ILEC and CLEC samples have at least 10 members. Thus, the issue of sample size should not generally be a problem.

25. There is an alternative method for developing the probability distribution of the test statistic that can be used with smaller sample sizes.[13] Under this method, called the permutation distribution, the probability distribution is generated through use of the actual sample results, rather than a preexisting table. Given two samples, X's and Y's from ILEC and CLEC respectively, we combine these into one pool and then divide this into two sets X^* and Y^* in all possible ways. For each way, we calculate the corresponding z-score, say z^*. This gives us a distribution of z^* values, each of which is equally likely under the null hypothesis that the ILEC is treating customers impartially. Given the desired Type I error rate, we can read off the appropriate critical value and compare this with the observed value.

26. For example, if the data are

3 ILEC observations: X=1, X=2, X=4
2 CLEC observations: Y=3 and Y=5

---

[13] This method will provide reliable results for any sample size, but the use of the t-distribution and the associated table is simpler for all but very small sample sizes.

then the pooled set is (1,2,3,4,5) and there are 10 ways we can assign these five observations to the ILEC and CLEC samples. We get 10 values of z:

-2.74 -1.20 -0.60 -0.44 0.00 0.00 0.44 0.60 1.20 2.74

and the 5% critical value is 2.74. The actual observed value is 1.20, and so is judged to be not significant (i.e., we accept the null hypothesis).

27. This test procedure is valid irrespective of the form of the population distribution, since it depends only on the assumption that each possible permutation is equally likely under the null hypothesis.[14] The method can be used whenever the sample sizes are large enough to make the test statistic well defined, in the present case even for m=2, n=1.

28. The permutation distribution would be developed through the use of a computer program that would enumerate the samples necessary to generate the distribution. I wrote a program to perform this function in a commercially available program language called S Plus in one-half hour. Thus, I believe that a suitable program could be developed

---

[14] See, e.g., Cox and Hinkley, Theoretical Statistics (1974) (paperback edition Chapman and Hall, 1979), pp. 182-184; H. Scheffe, The Analysis of Variance (1959) John Wiley & Sons, Section 9.3; P. Good, Permutation Tests (1994) Springer.

promptly for use by the entire ILEC industry at minimal cost[5].

29. A resource issue relating to the use of the permutation distribution is the time needed to generate results. Unless the sample sizes are very small, the number of permutations to be generated is extremely large.[6] In order to deal with this problem, it would be reasonable to use a random sample of possible permutations to approximate the distribution. For example, if the number of possible permutations in a particular case exceeds 1000, the program could be designed to approximate the permutation probability distribution by randomly selecting 1000 permutations and constructing the distribution from those data. Because computers can perform calculations such as this with remarkable speed, the distribution for any measurement category could be ascertained within a few seconds.[7]

---

[5] The Cytel Software Corporation of 675 Massachusetts Avenue, Cambridge, MA, markets a product called StatXact which has the capability of performing permutation tests.

[6] If m=10, n=5, there are 3003 permutations; if m=20, n=10, there are over 30 million.

[7] The Notice (Appx. B, n.5) raises another interesting possibility for a statistical analysis of individual performance measurements, i.e., comparing the proportions of two samples that exceed some fixed value. AT&T is studying a variation of this concept, in which the fixed value is not specified in advance, but is determined from the ILEC sample itself. We use the upper 90% quantile of the ILEC sample to determine the level of service that the ILEC is providing for 90% of its customers and then measure what percentage of CLEC customers receive at least that level of service. The

D. ILECs' Compliance With Their Nondiscrimination Obligations Should Be Based On An Aggregate Assessment Of Parity.

30. One of the key concepts in the AT&T Statistical Ex Parte is that it is also appropriate to use statistical analysis to review the aggregate results of an ILEC's performance to determine whether it is in compliance with its nondiscrimination obligations. If we apply a large number, several hundred perhaps, of tests of individual performance measurement comparisons, each test having a Type I error rate of 5%, then we would expect, on average, about 5% of these tests to indicate non-compliance even when the ILEC is actually fully in compliance. Thus the fact that this many tests indicate non-compliance does not give conclusive evidence that the ILEC is not in compliance with its Section 251 nondiscrimination obligations. The number of tests that erroneously indicate non-parity will vary randomly about this average number. We need to derive some

---

"parity" hypothesis is rejected if the fraction of CLEC customers receiving that level of service is much smaller than the percentage of ILEC customers receiving such service. (For example, if the ILEC completes repairs on a specific service for 90% of its customers within 48 hours, parity is not achieved if the ILEC complete repairs for much less than 90% of CLEC customers within that amount of time.) This test procedure is non-parametric, i.e., it does not require any assumptions beyond the basic one that under the null hypothesis CLECs receive equal treatment to the ILEC. This methodology only applies, however, to the review of individual performance tests. It does not address the need to develop a method to review ILEC performance in the aggregate.

threshold number of failed parity tests such that if more than this number are observed to fail, then non-compliance can be deduced. This threshold number of tests must be determined in such a way as to control the probability of an overall, or aggregate, Type I error at 5%. Furthermore, I also recommend that any review of an ILEC's compliance with its nondiscrimination obligation should be based on two dimensions of statistical comparisons, both of which must be satisfied.[11] The two dimensions of statistical comparisons are

(a) the number of tests that fail in any monthly period must not be too large, and

(b) the number of tests that fail for three consecutive months must not be too large.

Here, "too large" must be determined by consideration of the total number of individual tests and the desired overall Type I error rate.

31. For the first dimension, we must determine how many of the individual measurements subjected to the above comparison tests need to demonstrate non-parity before an ILEC may be found to be in overall violation of its

---

[11] The AT&T Statistical Ex Parte suggested that a third dimension also be considered, namely imposing a bound on the number of individual tests that exhibit extreme violations. I now judge that imposing this additional constraint does not provide much additional power for detecting extreme violations, and in fact reduces the chance of detecting some more moderate violations.

statutory duty. Suppose we have made N individual tests, each having a 5% Type I error rate, and have found that K of them indicate non-compliance. If K is approximately .05 times N, we have no conclusive evidence of overall non-compliance. Under the assumption that the ILEC is in compliance, we can determine a number k such that the probability that K exceeds k is 5%.[::]

32. The second dimension, i.e., the number of measurements failing the test repeatedly, is necessary to assure that the ILEC failures are indeed random. Without this dimension, the ILEC might be able to "game" the process and produce repeatedly discriminatory results on measures that are critical to one or more competitors. Thus, for this dimension, we must determine how many individual measurements in an ILEC report may be allowed to fail the parity test in three successive months before finding that the ILEC has failed to provide parity.

33. Suppose we have made N individual tests for each of three months, each test having a Type I error of 5%. Let K be the number of tests that have failed in all three months. The probability that any individual test fails in

---

[::] This computation assumes that under the null hypothesis, the number K has a binomial distribution with exponent N, i.e., it is as though we had tossed N coins, each with a probability of coming down "heads", and have counted how many "heads" appear. Then we claim non-compliance if K exceeds k.

all three months, assuming that the ILEC is in compliance with its nondiscrimination obligation, is $(.05)^3$, or 1/8000. Thus the expected number that fail in all three months, assuming compliance, is N/8000. Given that the number of monthly tests will be well below 8000, noncompliance should be found if K_2 is not zero. In other words, the allowed number of three-time-failing tests is $k_2=0$.

34. If we apply both of these overall procedures simultaneously, the actual overall Type I error rate is a function of three things: the Type I error rates of the individual tests, which I call α_1, the number k_1 of allowed individual failures, and the number k_2 of allowed three-time failures. These three numbers can be determined so that the Type I error rate of the overall procedure is exactly 5% (or whatever other value is required). Details of this computation are given in Exhibit 1.

II. BellSouth's Proposed Methodology Is Unsuited To Measure Parity And Should Be Rejected.

35. The Notice (Appx. B, ¶ 7) also solicits comments on the methodology proposed by BellSouth, which is based on the use of statistical process control. This approach is not suitable to measure parity between ILECs and CLECs and should be rejected.

36. BellSouth has proposed three kinds of control charts. In the first, described in the Notice (Appx. B,

¶ 6), BellSouth maintains its own monthly results (presumably for each type of measurement) on a control chart. Three-sigma limits are established by reference to BellSouth's historical record. Then, each month, results for the CLEC are plotted on the same chart, and parity is claimed if these values do not fall outside the limits.

37. A second proposal appears in BellSouth's Tennessee Section 271 proceeding (see memo from David Laney to William Stacy, attached to the rebuttal testimony of William N. Stacy, TRA Docket 97-00309, Exhibit WNSPM-2). Here the proposal is to plot values of the variable DIFF=(CLEC value - ILEC value) on a control chart, with limits set at +/- 2.66 times the average moving range of size two.

38. A third proposal also appears in the same document from BellSouth's Tennessee Section 271 proceeding. Here it is proposed to compute z-scores, but using the process standard deviation in the denominator rather than the within-month ILEC sample variance as AT&T recommends. This process standard deviation is the average moving range (presumably of size two) divided by 1.128.

39. Each of these proposals has serious deficiencies, the most serious being that statistical process control is not designed to measure differences in parity. Rather, this technique is used to measure stability in performance. Stability of ILEC processes is of course an important

concept, because the overall reliability of the systems used to serve CLECs is essential to determining whether an ILEC has met its duties under Section 251 of the Act. However, it is irrelevant in determining whether an ILEC's performance for itself is at parity with the performance it provides to others, i.e., CLECs. The ILEC's performance could be stable, with parity not provided, or unstable with parity being provided. Stability and parity are distinctly different concepts.

40. Another shortcoming of each of the three BellSouth proposals is that no allowance is made for the fact that the number of observations that contribute to each average may change from month to month. This makes the use of moving ranges invalid measurements of variability. Also, the number of observations in the CLEC sample is very unlikely to equal the number in the ILEC sample. Thus the ILEC and CLEC averages will not have the same variances, even assuming parity, and so should not be compared to the same control limits, as the first proposal suggests.

41. If control limits for the quantity DIFF were to be set using the process variability of this quantity, as in the second and third proposals, some consistent violations of parity could completely avoid detection. Namely, if for any reason the CLEC measurements were consistently more variable than the ILEC measurements (which would imply that

many CLEC customers were getting poorer service), then this variability would be included in setting the control limits, and lack of parity would not be detected.

42. Further, use of separate control charts for each of the many types of measurement leaves open the question of how an overall judgement of compliance should be arrived at. BellSouth has not addressed this issue.

Conclusion

43. In summary, my testimony shows that AT&T's proposed methodology satisfies the Commission's desire to assure that reported differences in ILEC performance are statistically meaningful.

44. With respect to individual tests of ILEC performance, there are three key components in developing an appropriate statistical methodology. First, the modified z-statistic proposed by LCUG provides an appropriate test statistic to determine whether there are significant differences in the mean and the variance of an ILEC's performance for itself and for CLECs. Second, a one-tailed test with Type I error at about the 5% level strikes a fair balance between the need to account for both Type I and Type II errors. Third, the t-distribution provides a useful basis for calculating the critical value for individual tests of ILEC performance. Moreover, in those few cases

where the size of the ILEC sample is small, use of the permutation distribution will provide valid results.

45. It is also appropriate to aggregate the results of individual tests to determine whether the ILEC is in overall compliance with its duty to provide nondiscriminatory treatment to CLECs. This should be done through the use of a two-part analysis that sets limits on the number of individual tests that fail to demonstrate parity in any given month and on the number of individual tests that fail in three consecutive months. These limits can be determined in such a way that the overall Type I error is held at 5%.

46. Finally, the methodology suggested by BellSouth is not designed to measure parity of performance between two different populations. Thus, it should not be used to determine whether ILECs have met their legal duty to provide CLECs with parity service.

Colin L. Mallows

Colin L. Mallows

Sworn to before me this
29th day of May, 1998

Patricia A. Parker
Notary Public

My Commission expires 4/8/2002

Exhibit 1

Statistical Definition of the Compliance Rule for ILEC Parity

The number k_1 of allowed individual violations, and the Type I error of each of the individual tests[22], α_1, are determined so that the probability of falsely claiming overall violation is controlled at a known level[23], which we call α.

Suppose we are aggregating N individual tests. Let K_1 be the number of these tests that indicate violations this month, and let K_2 be the number of tests that have shown violations in each of the past three months. Our proposed procedure is to claim overall violation if either (i) K_1 exceeds some number k_1, or (ii) K_2 exceeds zero. We show how k_1 and the type I error α_1 of each individual test can be determined so that the Type I error of the overall procedure is held at some desired level α.

To determine k_2 and α_2 when we know N, (the number of tests to be aggregated), and α, we proceed as follows. Throughout this calculation, we are assuming that the ILEC is fully in compliance, so that for each individual test the probability of (falsely) indicating non-parity is α_1.

a) Choose a tentative value for α_2. We start with $\alpha_1 = \alpha$. This value of α_1 will be adjusted downwards later.

b) Determine k_2 to be the largest number such that the probability that the overall procedure indicates violation[24] is greater than α.

c) Decrease α_1 until the probability of overall violation using the value of k_1 that was determined in step b), is exactly α.

---

[22] Also referred to as the size of the individual test.

[23] Also referred to as the size of the overall aggregated test.

[24] This probability is: $1 - (1 - \alpha_1^3)^N * P(k, N, p)$ where $P(,,)$ is the cumulative probability of the binomial distribution. That is, $P(k, N, p)$ is the probability that the number of false parity test failures is <= k when the probability of an individual false parity test failure is p, and where $p = (\alpha_1 - \alpha_1^3)/(1 - \alpha_1^3)$.

The resulting value of α_1 (and the corresponding critical value on the z-score scale) is to be used in each of the individual tests. Then non-compliance is indicated if any series fails the test in three successive months, or if more than k_1 fail in any single month.

The following table provides an example of how k_1 is determined for the values N = 100 and α = 5%. As shown, the value of k_1 = 8 is the largest value of k that corresponds to a probability of no less than 5% of being exceeded. In this case, the probability of claiming an overall violation is 7.40%.

Table 1

Determination of k_1 for N=100, α= 5%

k	Prob($K_1>k$, $K_2>0$) =1 - (1 - α_1^3)^N * P(k, N, p)	
5	38.95%	
6	24.17%	
7	13.76%	
8	7.40%	← select this k for k
9	3.99%	
10	2.36%	

The next step is to iteratively decrease α_1 and recompute the overall probability of violation, with k_1 held at 8, until we arrive at a value for α_1 for which this probability is .05. In this case, that value of α_1 is .04601.

Now we can use the t-tables (or permutation distribution calculations) to determine the appropriate critical values for each individual test. The following Table 2 provides k_1, α_1, and critical values (assuming large sample sizes for each test) for α = .05 and a number of values of N.

Table 2

Determination of k_1 and α_1 for a range of N
where k_1 satisfies $1 - (1 - \alpha_1')^N * P(k_1, N, p) = .05$

N	k_1	k_1 as a % of N	α_1	Critical Value (c)
70	6	8.57%	.0465	1.6803
80	[illegible]	[illegible]	.0408	1.7411
90	7	7.78%	.0437	1.7096
100	8	8.00%	.0460	1.6849
120	9	7.50%	.0442	1.7038
140	10	7.14%	.0430	1.7170
160	12	7.50%	.0462	1.6825
180	13	7.22%	.0452	1.6937
200	14	7.00%	.0443	1.7026
250	17	6.80%	.0441	1.7046
300	20	6.67%	.0440	1.7060
400	26	6.50%	.0437	1.7095
500	32	6.40%	.0431	1.7155
600	38	6.33%	.0423	1.7247
700	44	6.29%	.0412	1.7374
800	49	6.13%	.0397	1.7543
900	55	6.11%	.0384	1.7696
1000	60	6.00%	.0371	1.7851